Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering 15 (2011) 5563 – 5567
Procedia Engineering
www.elsevier.com/locate/procedia
Advanced in Control Engineering and Information Science

# Enhanced classification based on probabilistic extreme learning machine in wastewater treatment process

L.J. Zhao[a,b]*, X. K. Diao[a], D. C. Yuan[a], W. Tang[c]

[a] *Province Key Laboratory of Process Control Technology for Chemical Industry, Shenyang University of Chemical Technology, ,Shenyang, 11004, China*

[b]*State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University,Shenyang, 110004, China*

[c] *China Heating Company, Haicheng, 114200, China*

**Abstract**

A binary classification method, Probabilistic Extreme Learning Machine (called P-ELM), is proposed to enhance the reliability of the classification of an unknown object. P-ELM method integrates ELM, density methods and Bayesian decision theory in order to take into account a priori probability of the process and the uncertainty of the ELM predictions. The P-ELM algorithm may inhibit uncertainty of the extreme learning machine prediction in the different trials of simulation due to the initialization of input weights and bias, which would damage the reliability of the classification for the new objects. Simulations results from a municipal wastewater treatment plant show that the proposed P-ELM binary classification method can provide the reliability of the classification for those samples near the boundaries of the classes and the reliability and accuracy outperform the ELM model.



Keywords: ELM; probability ELM; binary classification; Bayesian decision;

## 1. introduction

Reliable classification of operational states is necessary for the optimization and decision in wastewater treatment plants. Due to the complexity of the wastewater treatment processes, existing control technology has not been applied effectively, which results in the big fluctuation of effluent

* Corresponding author. Tel.: +86-024-89383907
*E-mail address*: zlj_lunlun@163.com.


doi:10.1016/j.proeng.2011.08.1032

quality, low efficiency and high cost. Operational condition monitoring is essentially a pattern classification problem, so it needs pattern classification method to achieve process operation state monitoring [1].

The classification of operational states by use of a combination of PCA and FCM clustering has been described [2,3]. The approaches are based on the fact that different operational states (caused by disturbances) generally manifest themselves as clusters. Clustering representing normal operation as well as different disturbance types is defined in the reduced space in order to reduce the dimensionality and to decrease the noise level. These methods are unsupervised learning completed without the presence of the labelled patterns. Discriminant partial least squares is a supervised learning with the available prior knowledge and provides the reliability of the classification integrating density method and Bayes decision theory [4]. Extreme Learning Machine (ELM) tends to provide good generalization performance at extremely fast learning speed and has been effectively used in classification applications [5,6]. But ELM algorithm may have uncertainty in different trials of prediction due to the initialization of input weights and bias, which would damage the reliability of the classification for the raw data. In the study, ELM is integrated with density methods and Bayesian decision theory in order to take into account the uncertainty of the predictions in ELM. Simulations results from a municipal wastewater treatment plant show that the proposed P-ELM binary classification method can provide the reliability of the classification.

## 2. Theory and methods

The proposed P-ELM binary classification method is as follows:

### *2.1 Extreme learning machine Model*

For $I$ arbitrary distinct samples $\{x_i, y_i\}, i = 1, \cdots, I$, Y-block is firstly coded with the integer 1 if the sample belongs to the class of interest (class $\omega_1$) or 0 otherwise (class $\omega_0$). ELM model is performed on a training set, $X(I \times J)$, and an indicator matrix representing group membership $Y(I \times 2)$. Given hidden node output function $G(a, b, \mathrm{x})$, and the number of hidden nodes $L$, ELM algorithm is as follows:

(1) Assign randomly hidden node parameters $(a_i, b_i), i = 1, \cdots, L$.

(2) Calculate the hidden layer output matrix $H$.

(3) Calculate the output weight $\beta_i$: $\hat{\beta} = H^{\dagger} Y$.

where $H^{\dagger}$ is the Moore-Penrose generalized inverse of hidden layer output matrix $H$.

### *2.2 Calibration and probability density functions for classes in ELM*

ELM prediction $\hat{Y}$ for a calibration set is:

$$\widehat{Y} = H\beta \tag{1}$$

where $\boldsymbol{H} = \begin{bmatrix} g(\boldsymbol{a}_1 \cdot \boldsymbol{x}_1 + b_1) & \ldots & g(\boldsymbol{a}_{\mathrm{L}} \cdot x_1 + b_{\mathrm{L}}) \\ \ldots & \ldots & \ldots \\ g(\boldsymbol{a}_1 \cdot \boldsymbol{x}_{\mathrm{N}} + b_1) & \ldots & g(\boldsymbol{a}_{\mathrm{L}} \cdot \boldsymbol{x}_{\mathrm{N}} + b_{\mathrm{L}}) \end{bmatrix}_{\mathrm{N} \times \mathrm{L}}$, $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \ldots \\ \boldsymbol{\beta}_L^T \end{bmatrix}_{L \times 2}$ and $\hat{\mathbf{Y}} = \begin{bmatrix} \hat{\boldsymbol{y}}_1^T \\ \ldots \\ \hat{\boldsymbol{y}}_I^T \end{bmatrix}_{I \times 2}$.

The standard error of prediction (SEP) is used to account for the prediction uncertainty of the ELM model. $SEP_i$ for each calibration sample $i$ is calculated by:

$$SEP_i = \sqrt{(1+h_i)\times MSEC_{bc}} \tag{2}$$

where $h_i$ is the leverage for sample $i$ and $MSEC_{bc}$ is the bias corrected mean squared error of calibration. The Leverage value is calculated by $h_i = \mathrm{x}_i^T(\mathrm{X}^T\mathrm{X})^{-1}\mathrm{x}_i$. The bias-corrected $MSEC_{bc}$ is calculated as:

$$MSEC_{bc} = \frac{1}{I-J-2}\sum_{i=1}^{I}(\hat{y}_i - y_i - bias_c)^2 \tag{3}$$

where $bias_c = \frac{1}{I_c}\sum_{i=1}^{I_c}(\hat{y}_i - y_i)^2$ is the bias of class $c\left(c=0,1\right)$. The potential functions of the training samples for class $c$ are averaged to obtain the PDF of class $\omega_c, c=0,1$:

$$p(\hat{y}|\omega_c) = \frac{1}{I_c}\sum_{i=1}^{I_C} g_i(\hat{y}), c=0,1 \tag{4}$$

where $g_i(\hat{y})$ is probability density function of each calibration sample $i$ for classes $\omega_0$ and $\omega_1$ with the shape of a Gaussian curve, centred at $\hat{y}_i$ and standard deviation $SEP_i$

$$g_i(\hat{y}) = \frac{1}{SEP_i\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{\hat{y}-\hat{y}_i}{SEP_i})^2} \tag{5}$$

*2.3 Prediction and classification*

Suppose that the prior probabilities $p(\omega_c) = I_c/I$ and the conditional densities $p(y\mid\omega_c)$ for $c=0,1$. For an unknown testing sample, the prediction $\hat{y}_u$ and its standard error of prediction $SEP_u$ are calculated following Eqs. (1)-(2). The probability with prediction $\hat{y}_u$ belongs to the class $\omega_c$ is given by the Bayes formula:

$$p(\omega_c|\hat{y}_u) = \frac{p(\hat{y}_u|\omega_c)\times p(\omega_c)}{p(\hat{y}_u|\omega_0)\times p(\omega_0) + p(\hat{y}_u|\omega_1)\times p(\omega_1)} \tag{6}$$

Bayes formula shows that the prior probability $p(\omega_c)$ is converted into the a posteriori probability $p(\omega_c|\hat{y}_u)$ by prediction $\hat{y}_u$ when the prediction takes well-defined values. However, the true value is within an interval $U_u = \{\hat{y}_{u,1} \le \hat{y}_u \le \hat{y}_{u,r}\}$ . $\hat{y}_{u,1} = \hat{y}_u - k\cdot SEP_u$ and $\hat{y}_{u,r} = \hat{y}_u + k\cdot SEP_u$ (k being a coverage factor $k=1$ and $k=2$), are the left and right limits of the interval . Bayesian decision formula for assigning an unknown testing sample is rewritten as [4]:

$$p\{\omega_c|\hat{y}_{u,1} \le \hat{y}_u \le \hat{y}_{u,r}\} = \frac{p\{\hat{y}_{u,1} \le \hat{y}_u \le \hat{y}_{u,r}|\omega_c\}}{p\{\hat{y}_{u,1} \le \hat{y}_u \le \hat{y}_{u,r}\}}p(\omega_c) = \frac{\int_{\hat{y}_{u,1}}^{\hat{y}_{u,r}} p(\hat{y}_u|\omega_c)d\hat{y}_u}{\int_{\hat{y}_{u,1}}^{\hat{y}_{u,r}} p(\hat{y}_u)d\hat{y}_u}p(\omega_c) \tag{7}$$

The numerator of Eq. (7) is the area under the curve $p(\hat{y}_u|\omega_c)\times p(\omega_c)$ in the interval $\{\hat{y}_{u,1} \le \hat{y}_u \le \hat{y}_{u,r}\}$:

$$Area_{u,c} = p(\omega_c)\int_{\hat{y}_{u,1}}^{\hat{y}_{u,r}} p(\hat{y}_u|\omega_c)d\hat{y}_u \text{ , } c=0,1 \tag{8}$$

Since the denominator in Eq. (8) is not fundamental for the decision, the rule can be expressed as:

Decide class $\omega_1$ if $Area_{u,1} > Area_{u,0}$; otherwise decide class $\omega_0$. (9)

## 3. Results and Discussion

### *3.1. Description of wastewater treatment process*

The case study is a small-scale wastewater treatment plant located in Liaoning province. The schematic diagram of the process is shown in Fig. 1. The plant has the daily capacity of about 3 million $m^3$, and COD in the influent about 360g/ $m^3$. Total hydraulic retention time is about 19 hours, and sludge age about 12 days. Table 1 lists on-line variables, sample time 1h.

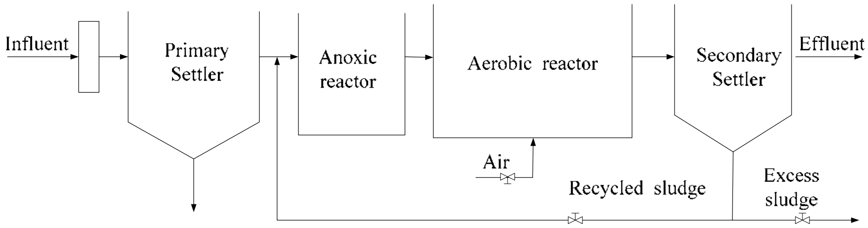


Fig.1 Schematic diagram of the activated sludge wastewater treatment process

Table 1. Variables in the P-ELM Model

| Symbol | Unit | Description | Symbol | Unit | Description |
|---|---|---|---|---|---|
| $Q_{in}$ | $m^3/h$ | Volumetric flow rate in the influent | $Q_{eff}$ | $m^3/h$ | Volumetric flow rate in the effluent |
| $COD_{in}$ | mg/L | COD in the influent | $COD_{eff}$ | mg/L | COD in the effluent |
| $MLSS_1$ | mg/L | 1# Sludge concentration in the anoxic tank | $MLSS_2$ | mg/L | 2# Sludge concentration in the aeration tank |
| $DO_1$ | mg/L | 1# Dissolved oxygen concentration in the anoxic tank | $DO_2$ | mg/L | 2# Dissolved oxygen concentration in the aeration tank |
| $pH_{in}$ | — | PH in the influent | | | |

### *3.2. Comparisons between ELM and P-ELM model*

For WWTP dataset, a model was established to classify the samples the class $\omega_1$ or not (class $\omega_0$). $\omega_0$ denotes normal operational states, and $\omega_1$ denotes overload operational states. In this case of two categories, ELM and P-ELM classification performance were compared. Firstly, each class was labelled according to the expert priori knowledge. A training set with the 600 samples and a test set with 150 samples were used. Fig. 2 shows the probability density function, the probability density function multiplied by the a priori probability, and the a posteriori probability for the classes $\omega_0$ and $\omega_1$.

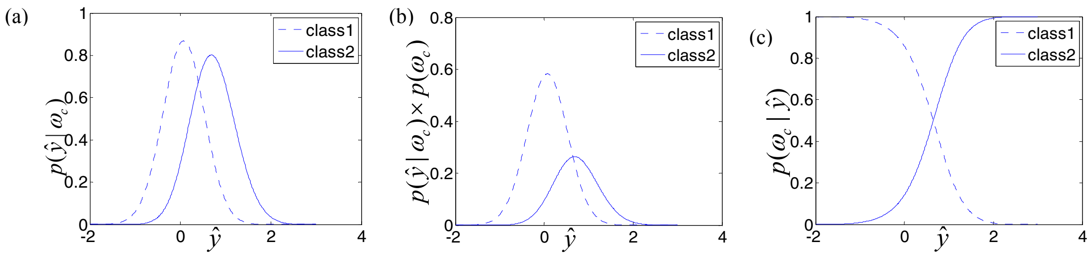


Fig. 2. (a)PDF of two classes; (b) PDF multiplied by the prior probability of the class ;( c) posterior probabilities

Fig. 2 shows ELM and P-ELM prediction results for the testing sample. Classification results of ELM and

P-ELM were shown in Table 2. Observed from Table 2, testing accuracy in P-ELM was more than the testing accuracy in ELM.

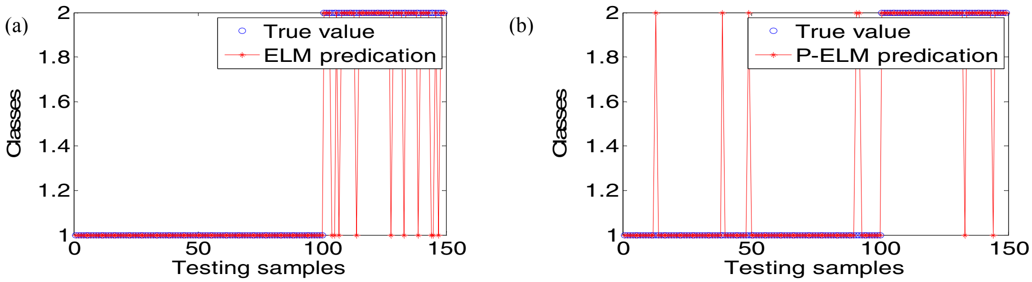


Fig. 3. (a) ELM result of testing set; (b) P-ELM result of testing set

Table 2. Performance comparison between ELM and P-ELM

| | ELM | P-ELM |
|---|---|---|
| Training | 95.0847% | 93.2886% |
| Testing | 95.2542% | 95.3025% |

## 4. Conclusions

Probabilistic extreme learning machine is proposed to enhance reliability of classification through integrating ELM with Bayesian decision theory. PDF for binary classes have been constructed by averaging individual kernel functions centered in the predictions of the training set for each class. The reliability of the classification was calculated from the area under the curve within two limits defined by the standard error of prediction. Simulations results from a municipal wastewater treatment plant show that the reliability and accuracy of the probabilistic extreme learning machine outperform the ELM model.

## Acknowledgments:

The work is supported by NSF in China (61020106003 and 60874057) and China's Postdoctoral Science Foundation (20100471464).

## References

[1] Rosen C. , Yuan Z. Supervisory control of wastewater treatment plants by combining principal component analysis and fuzzy c-means clustering. Water Science Technology 2001; **43(7)**: 147-156.

[2] Pekka Teppola, Satu-Pia Mujunen and Pentti Minkkinen. Adaptive Fuzzy C-Means clustering in process monitoring. *Chemometrics and Intelligent Laboratory Systems* 1999; **45(1-2):** 23-38.

[3] Rosana K. Tomita, Song W. Park, Oscar A. Z. Sotomayor. Analysis of activated sludge process using multivariate statistical tools-a PCA approach. *Chemical Engineering Journal* 2002; **90(3):** 283-290.

[4] Néstor F. Pérez, Joan Ferré, Ricard Boqué. Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chemometrics and Intelligent Laboratory Systems* 2009; **95:**122-128.

[5] Huang G B, Zhu Q Y,Siew C K. Extreme learning machine:Theory and applications. *Neurocomputing* 2006,**70(1-3)**:489-501.

[6] Huang G. B. and Chen L. Enhanced random search based incremental extreme learning machine, *Neurocomputing* 2008, **71:** 3060-3068.